02/19/25 - Graph Data Model

Graph Database

Data model based on the graph data structure

Composed of nodes and edges

- Edges connect nodes
- Each is uniquely identified
- Each can contain properties (e.g., name, occupation, etc.)
- Supports queries based on graph-oriented operations
    - Traversals
    - Shortest path
    - And more

Where do graphs show up?

Social networks

- Modeling social interactions in fields like psychology and sociology

The web

- It’s just a big graph of pages (nodes) connected by hyperlinks (edges)

Chemical and biological data

Graphs of Graph Theory

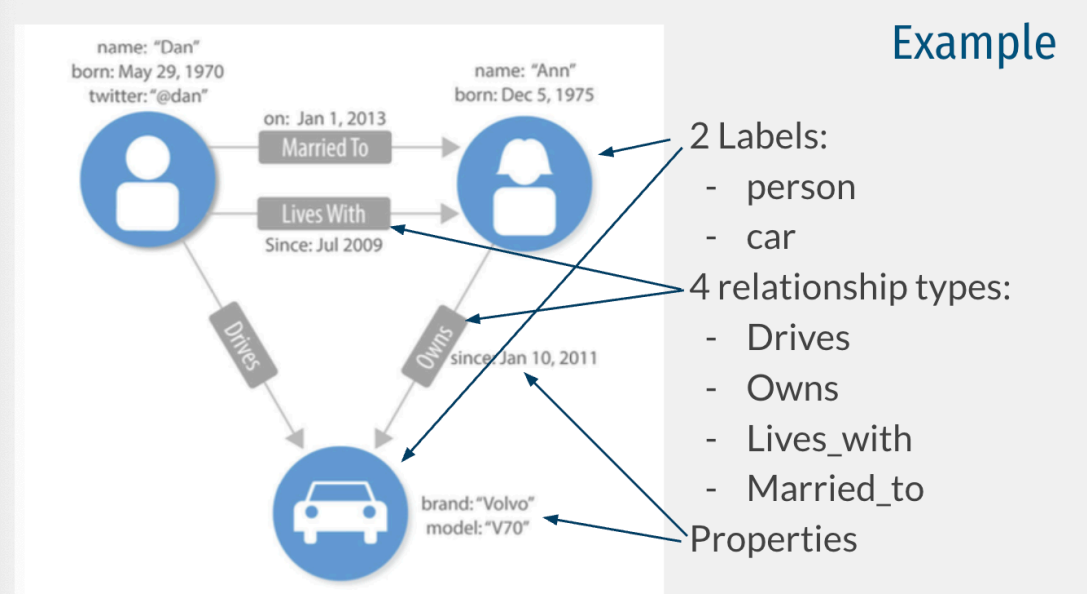
Labeled Property Graph

Composed of a set of node (vertex) objects and relationship (edge) objects

- Nodes with no associated relationships are ok
- Edges not connected to nodes are not permitted

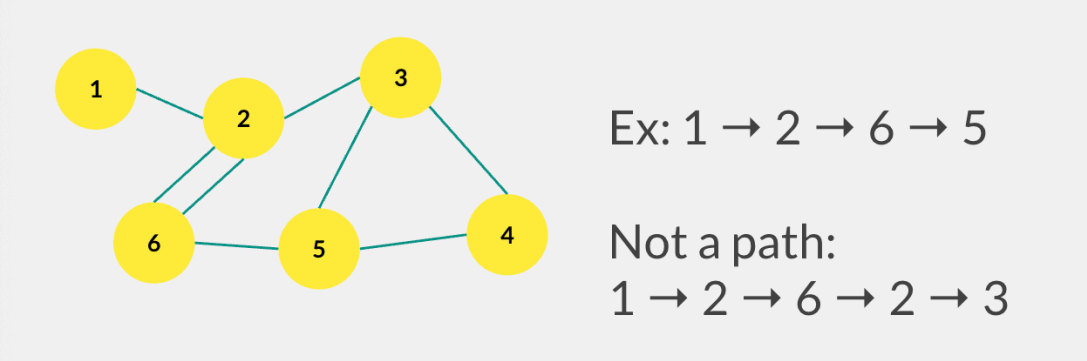
Labels are used to mark a node as part of a group

Properties are attributes (think KV pairs) and can exist on nodes and relationships



Paths

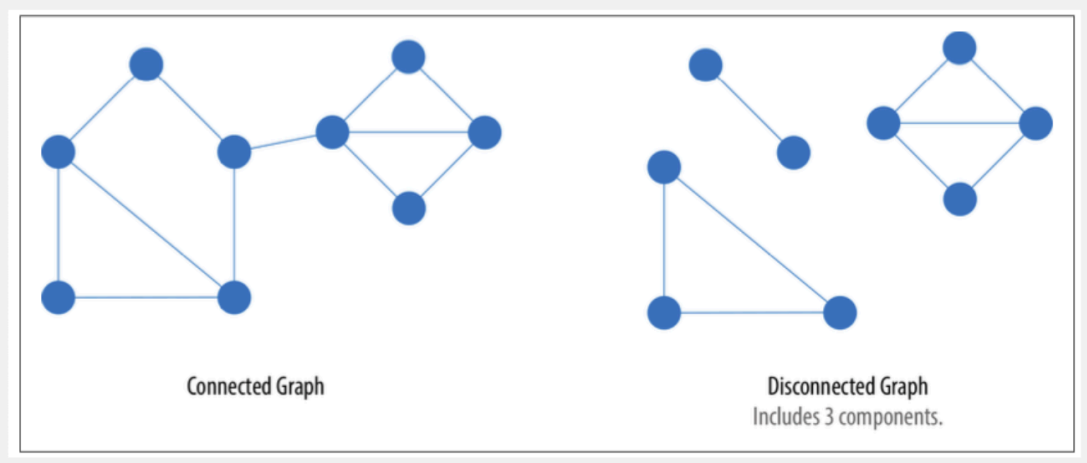
An ordered sequence of nodes connected by edges in which no nodes or edges are repeated



Flavors of Graphs

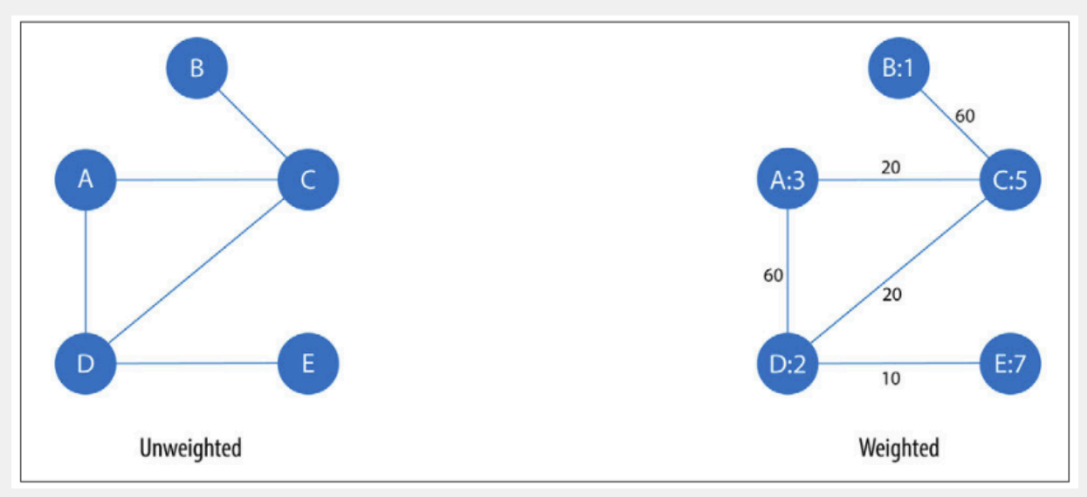
Connected vs Disconnected

There is a path between any two nodes in the graph



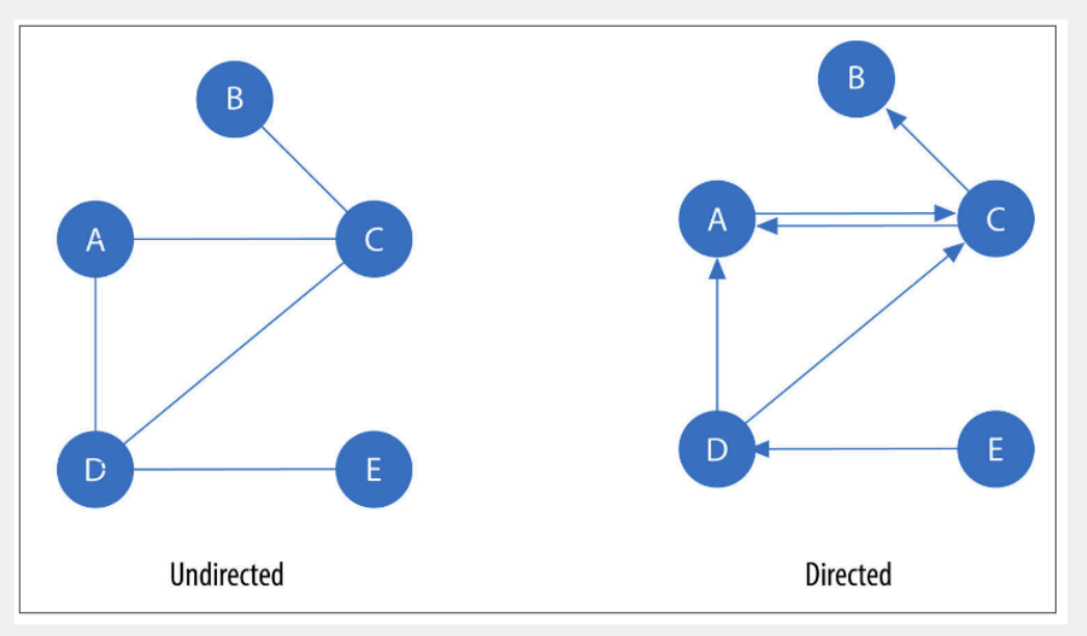
Weighted vs Unweighted

Edge has a weight property (important for some algorithms)



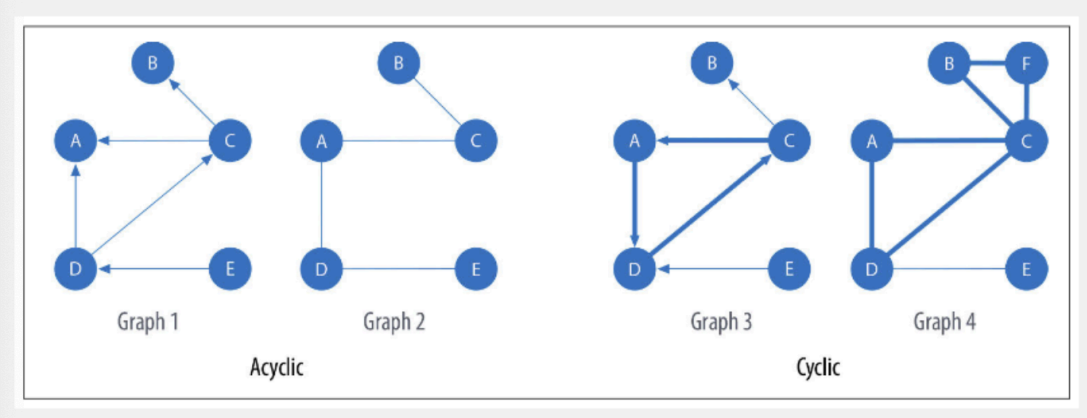
Directed vs Undirected

Relationships (edges) define a start and end node

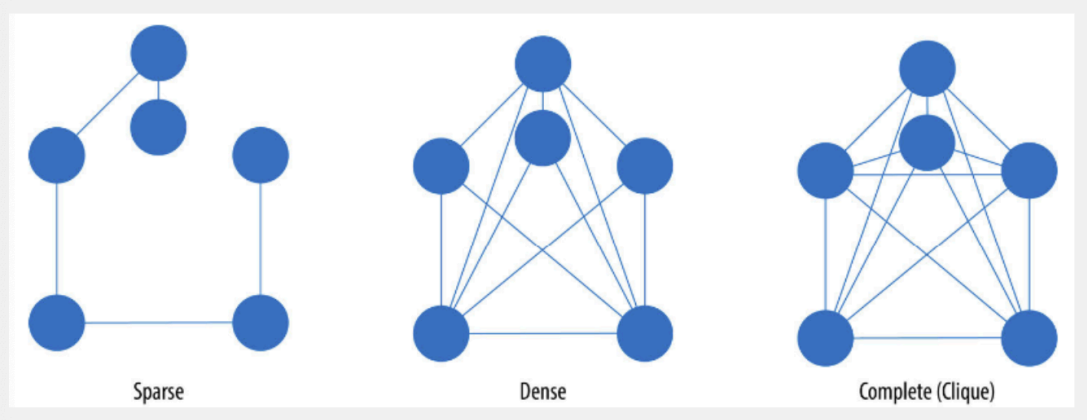


Acyclic vs Cyclic

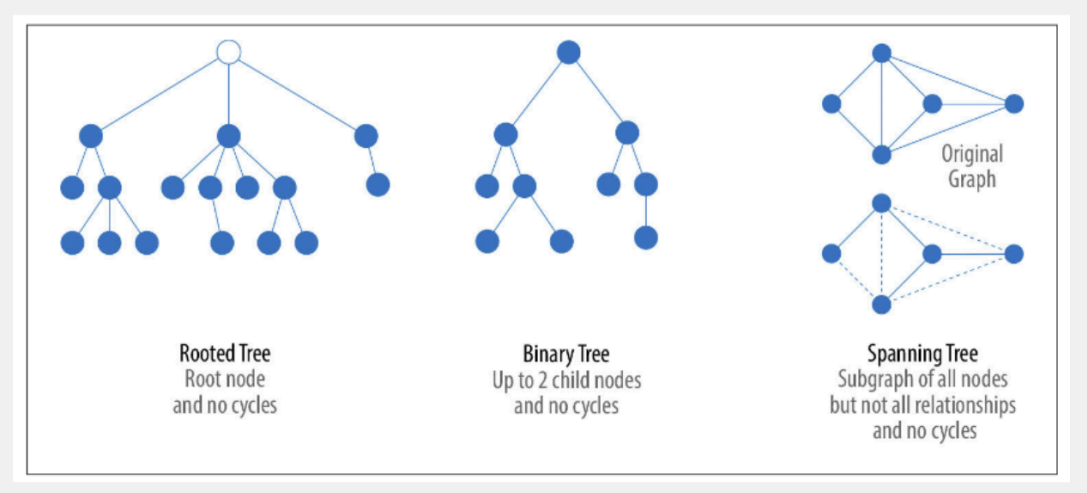
Graph contains no cycles



Sparse vs Dense



Trees



Types of Graph Algorithms

Pathfinding

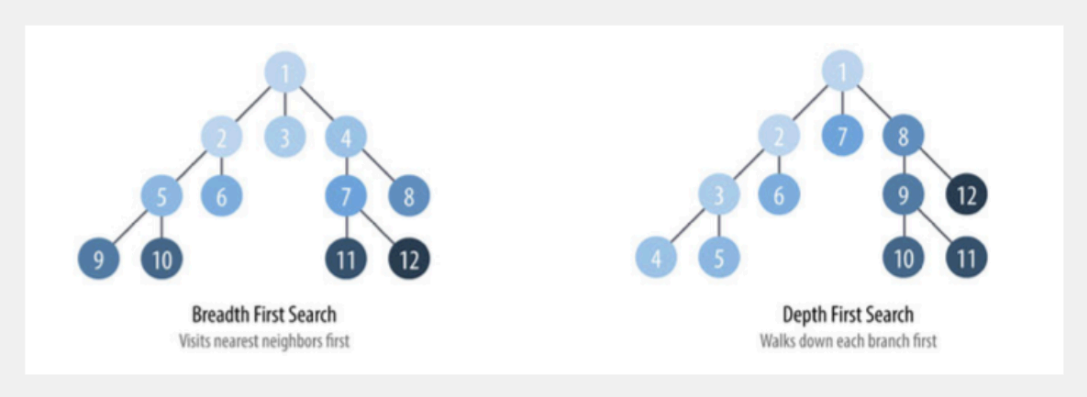
Finding the shortest path between two nodes, if one exists, is probably the most common operation

- Shortest means fewest edges or lowest weight

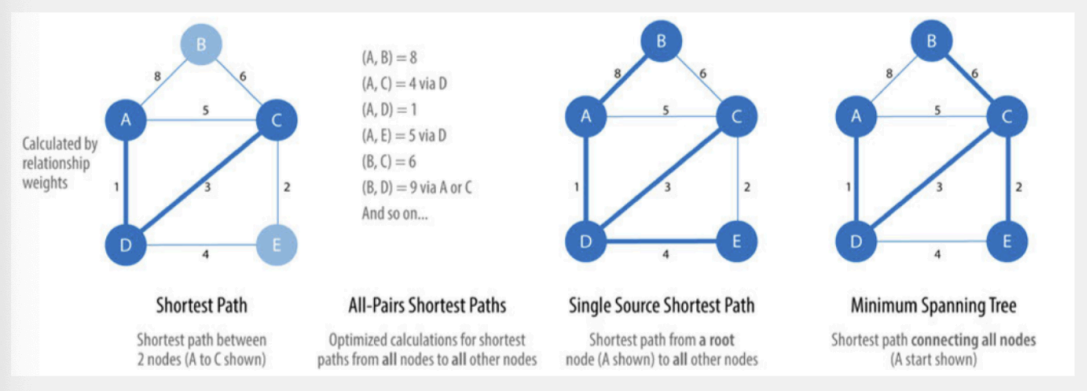
Average shortest path can be used to monitor efficiency and resiliency of networks

Minimum spanning tree, cycle detection, max/min flow, ... are other types of pathfinding

BFS vs DFS

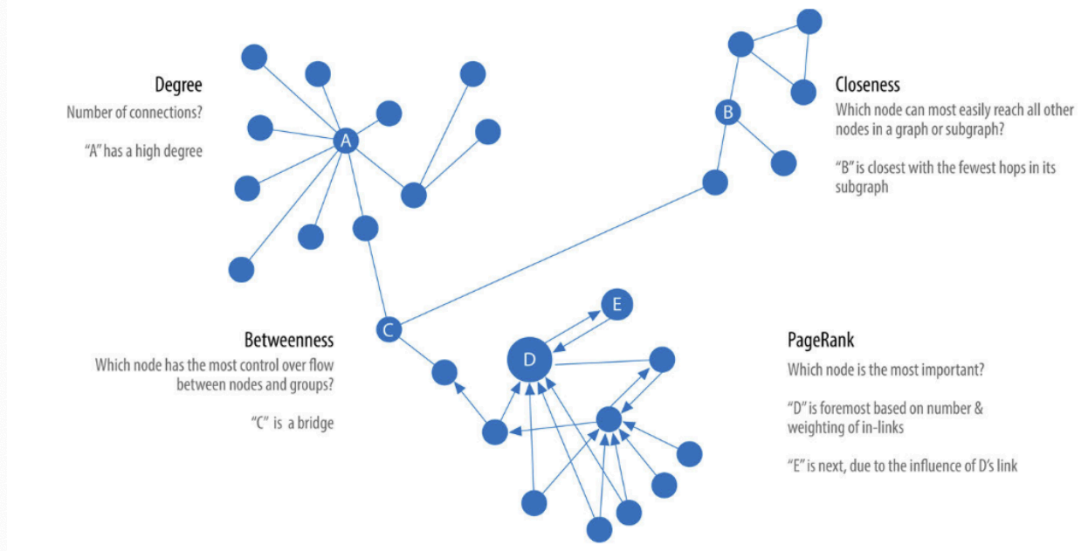


Shortest Path



Centrality

Determining which nodes are “more important” in a network compared to other nodes



Community Detection

Evaluating clustering or partitioning of nodes of a graph and tendency to strengthen or break apart

Some Famous Graph Algorithms

Dijkstra's algorithm: single-source shortest path algorithm for positively weighted graphs

A* algorithm: similar to Dijkstra's with added feature of using a heuristic to guide traversal

PageRank: measures the importance of each node within a graph based on the number of incoming relationships and the importance of the nodes from those incoming relationships

Neo4j

Graph database system that supports both transactional and analytical processing of graph-based data

- Similar: Microsoft CosmoDB, Amazon Neptune

Relatively new class of NoSQL DBs

Considered schema optional (one can be imposed)

Supports various types of indexing

ACID compliant

Supports distributed computing